

## مقایسه کارایی مدل‌های درختی M5 و برنامه‌ریزی ژنتیک

### در تخمین بار رسوب معلق رودخانه‌ها

- ❖ **جواد ظهیری\***: استادیار سازه‌های آبی، گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان، ایران.
- ❖ **سیده حدیث شاهرخی**: دانش‌آموخته کارشناسی ارشد سازه‌های آبی، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان.
- ❖ **احمد جعفری**: استادیار سازه‌های آبی، گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان، ایران.

#### چکیده

برآورد بار رسوب معلق یکی از مهم‌ترین و پیچیده‌ترین بحث‌های هیدرولیک رسوب و مهندسی رودخانه به حساب می‌آید. بر همین اساس روش‌های متعددی جهت برآورد بار معلق رسوب ارائه شده که با توجه به تجربی بودن این روش‌ها، دقت پایینی داشته و نتایج روش‌های مختلف تفاوت زیادی با یکدیگر دارند. امروزه با پیشرفت علوم رایانه‌ای الگوریتم‌های متنوعی ابداع شده که از آن جمله می‌توان به الگوریتم‌های درختی اشاره کرد. در تحقیق حاضر ضمن بررسی روش منحنی سنج‌رسوب، از الگوریتم درختی M5 و برنامه‌ریزی ژنتیک (GP) به عنوان روش‌های نوین جهت برآورد بار معلق رسوب استفاده شده است. اطلاعات مورد استفاده شامل دبی جریان آب و دبی رسوب مربوط به پنج ایستگاه آب‌سنجی بهبهان و چمن‌نظام بر روی رودخانه مارون، جوکنک بر روی رودخانه الله و مشراکه و شادگان بر روی رودخانه جراحی در استان خوزستان است. در تمامی ایستگاه‌ها دقت دو مدل درختی از منحنی سنج‌رسوب بیشتر بوده و مقدار مجموع مربعات خطا در تمامی ایستگاه‌ها جهت مدل M5 به میزان ۷ الی ۴۱ درصد نسبت به منحنی سنج‌رسوب کمتر محاسبه شده است. علاوه بر این، نتایج این تحقیق نشان‌دهنده نزدیکی کارایی دو روش M5 و GP است که با توجه به ساختار ساده و مفهومی مدل M5 این روش به عنوان روش مناسب جهت برآورد بار معلق رسوب انتخاب گردید.

**کلید واژگان:** بار رسوب معلق، الگوریتم‌های درختی، منحنی سنج‌رسوب، الگوریتم M5، برنامه‌ریزی ژنتیک

## ۱. مقدمه

پدیده فرسایش خاک و انتقال رسوب از جمله فرآیندهای مهم هیدرودینامیک هستند که به عنوان یکی از بزرگ‌ترین چالش‌های بهره‌برداری از منابع آب‌های سطحی در جهان شناخته می‌شوند [۲]. تخمین مقدار بار رسوبی یا میزان انتقال آن برای بسیاری از پروژه‌های منابع آب از قبیل طراحی سدها و حوضچه‌های ذخیره آب سطحی، انتقال رسوب و آلودگی در رودخانه‌ها و دریاچه‌ها و نیز طراحی کانال‌ها و نگهداری آنها حیاتی است [۱۵]. از دیگر دلایل اهمیت تخمین رسوب معلق رودخانه‌ها می‌توان به تأثیر آن بر روی کیفیت سامانه‌های آبی اشاره کرد [۳]. در سال‌های اخیر استفاده از الگوریتم‌های هوشمند متفاوت در شاخه‌های مختلف مهندسی از جمله مهندسی رودخانه رو به افزایش بوده است. الگوریتم‌های فرا کاوشی توانایی تخمین پدیده‌ها و فرآیندهای پیچیده طبیعی را دارا بوده و اغلب دارای دقت بیشتری نسبت به روابط تجربی می‌باشند. در تحقیقی که از مدل درختی M5 جهت شبیه‌سازی ارتفاع امواج رودخانه استفاده گردید، به بررسی کارایی این الگوریتم در مقایسه با شبکه عصبی مصنوعی پرداخته شد. نتایج این مقایسه نشان داد که مدل درختی با دقت بیشتری می‌تواند ارتفاع امواج را شبیه‌سازی کند [۵]. همچنین در مطالعه‌ای که بر روی مدل‌های درختی CART و M5 جهت تخمین عمق آبشستگی اطراف پایه‌های پل صورت گرفت، مشخص شد که الگوریتم‌های درختی می‌توانند با دقت بالاتری نسبت به روش‌های موجود میزان آبشستگی را محاسبه نمایند [۱۰]. در تحقیقی دیگر جهت بررسی کارایی الگوریتم M5 در محاسبه عمق آبشستگی اطراف مجموعه پایه‌های پل از اطلاعات آزمایشگاهی و میدانی استفاده گردید. نتیجه این تحقیق دقت بالاتر مدل درختی در مقایسه با سایر روش‌ها بوده است [۴]. علاوه بر این از مدل‌های درختی M5 و CART جهت تخمین عمق آبشستگی اطراف پایه پل استفاده گردید. نتایج این تحقیق نشان داد که M5 با ارائه روابط رگرسیونی ساده می‌تواند میزان آبشستگی را با دقت

بیشتر در مقایسه با مدل درختی CART و سایر روش‌های متداول محاسبه نماید [۱۶]. شبکه‌های عصبی مصنوعی<sup>۱</sup> و ماشین بردار پشتیبان<sup>۲</sup> از جمله مدل‌های داده‌کاوی هستند که در زمینه برآورد رسوب معلق به طور گسترده مورد استفاده قرار گرفته‌اند. در تحقیقی که بر روی رودخانه دوبرج صورت گرفت، از دبی جریان روزانه و میزان بارندگی به عنوان پارامترهای ورودی استفاده شد. نتایج این تحقیق نشان داد که شبکه‌های عصبی مصنوعی و ماشین بردار پشتیبان می‌توانند با دقت بیشتر نسبت به روش‌های رگرسیونی میزان رسوب معلق را برآورد نمایند. علاوه بر این در این تحقیق از روش M-Test جهت تعیین داده‌های مورد نیاز آموزش مدل‌های داده‌کاوی استفاده گردید [۸]. در تحقیق دیگر که در حوضه حیدرآباد در غرب ایران انجام شد، از مدل‌های رگرسیون درختی<sup>۳</sup> جهت برآورد رسوب معلق استفاده گردید. متغیرهای مورد استفاده در این تحقیق شامل دبی جریان، مجموع دبی سه روز و مجموع بارندگی پنج روز متوالی بوده است. مقایسه نتایج حاصل از مدل‌های درختی با مدل‌های شبکه‌های عصبی مصنوعی و منحنی سنجه رسوب نشان‌دهنده دقت بالاتر مدل‌های درختی است [۱۲]. در مطالعه‌ای که بر روی میزان رسوب ورودی به رودخانه زرد صورت گرفت از اطلاعات مربوط به شش زیرحوضه رودخانه بایلو<sup>۴</sup> که یکی از مهمترین منابع تأمین رسوب رودخانه زرد به حساب می‌آید، استفاده گردید. در این تحقیق جهت تعیین روند تغییرات رسوب سالیانه از تست‌های Mann-Kendall و Pettitt استفاده گردید. براساس نتایج به دست آمده، طرح‌های صورت گرفته در زمینه حفاظت خاک در حوضه رودخانه زرد باعث کاهش رسوب ورودی به رودخانه گردیده است [۱۷]. از دیگر مطالعات صورت گرفته در زمینه کاربرد مدل‌های داده‌کاوی در برآورد رسوب معلق می‌توان به تحقیق صورت گرفته بر روی رودخانه کرو<sup>۱</sup> در

<sup>1</sup> Artificial Neural Networks (ANN)

<sup>2</sup> Support Vector Machines

<sup>3</sup> Regression Trees

<sup>4</sup> Beiluo River

اطلاعات و داده‌ها به صورت میدانی برداشت شده و شامل دبی جریان آب و دبی رسوب در پنج ایستگاه آب‌سنجی بهبهان و چمن‌نظام بر روی رودخانه مارون، جوکنک بر روی رودخانه الله و مشراگه و شادگان بر روی رودخانه جراحی در استان خوزستان است. مشخصات ایستگاه‌های مورد استفاده در جدول (۱) و موقعیت حوضه آبریز این ایستگاه‌ها در شکل (۱) نشان داده شده است. رودخانه مارون از ارتفاعات شرقی و شمالی حوضه آبریز و در حد فاصل رودخانه‌های کارون و زهره سرچشمه گرفته است. این حوضه عمدتاً استان‌های کهگیلویه و بویراحمد و خوزستان را در بر می‌گیرد. رودخانه جراحی از اتصال رودهای مارون و الله تشکیل گردیده است. این حوضه عمدتاً استان‌های خوزستان و کهگیلویه و بویراحمد و بخش کوچکی از استان چهارمحال بختیاری را در بر می‌گیرد. رودخانه الله نیز از منطقه‌ای کوهستانی واقع در ۷۰ کیلومتری شرق و شمال شرقی شهرستان رامهرمز سرچشمه گرفته و رامهرمز هم نامیده می‌شود.

الگوریتم M5 اولین بار توسط کوئینلان<sup>۳</sup> ابداع شد و پس از آن توسط ونگ و ویتن<sup>۴</sup> توسعه و بهبود یافت [۱۱] و [۱۳]. روش M5 شاخه‌های خود را به صورت دوتایی و تنها بر اساس یک متغیر ایجاد می‌کند، بدین گونه که بر اساس شرطی که در هر گره تعریف می‌شود، اطلاعات در آن گره به دو قسمت تقسیم می‌شوند [۷].

ترکیه اشاره کرد. در این تحقیق از شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان و منطق عصبی فازی تطبیقی<sup>۲</sup> استفاده گردید. نتایج این مطالعه نشان داد که مدل بردار پشتیبان با ورودی‌های دبی همان روز، دبی روز قبل و بار رسوب روز قبل، نسبت به بقیه مدل‌ها از کارایی بالاتری برخوردار است [۱].

در تحقیق حاضر جهت محاسبه بار معلق رسوب از دو مدل M5 و GP با متغیرهای ورودی دبی جریان همان روز و دبی جریان روز قبل استفاده شده است. مدل درختی M5 با توجه به ساختار ساده قادر به تقسیم فضای مسئله به چند دامنه و ارائه رابطه برای هر زیردامنه می‌باشد، در حالیکه مدل GP از ساختار درختی جهت بهینه‌سازی حل مسئله استفاده می‌کند. جهت بررسی کارایی مدل‌های درختی در این تحقیق از اطلاعات پنج ایستگاه آب‌سنجی استفاده گردید و در نهایت نتایج حاصل با مقادیر رسوب محاسبه شده توسط روش منحنی سنجه رسوب که متداول‌ترین روش برآورد رسوب در رودخانه‌ها به حساب می‌آید، مقایسه گردید.

## ۲. روش شناسی تحقیق

اطلاعات و داده‌های مورد استفاده جهت ساخت و صحت سنجی دو مدل درختی M5 و GP از منابع سازمان آب و برق استان خوزستان جمع‌آوری گردید. این

جدول ۱. مشخصات ایستگاه‌های آب سنجی مورد استفاده در این مطالعه

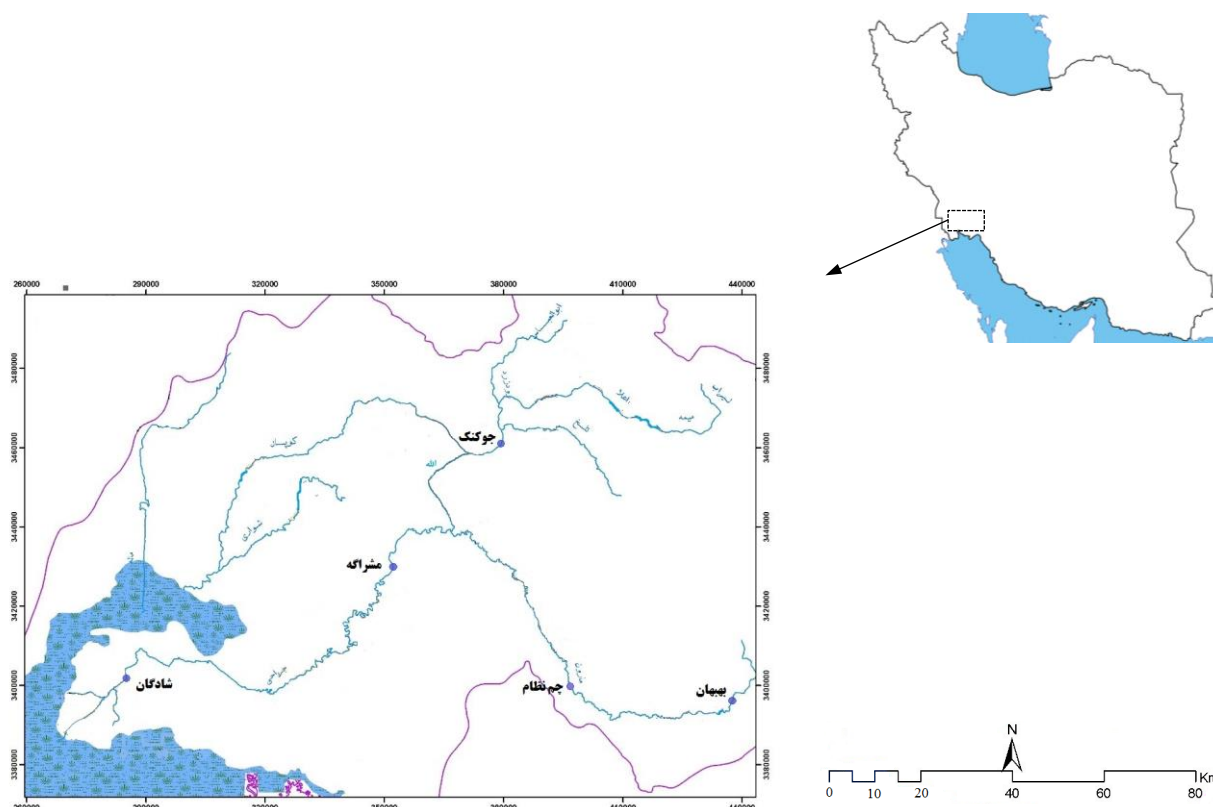
ردیف	ایستگاه	دوره آماری	موقعیت		
			عرض جغرافیایی (درجه)	طول جغرافیایی (درجه)	ارتفاع (متر)
۱	مشراگه	۱۳۶۶-۱۳۹۰	۳۱° ۰۰' ۲۲''	۴۹° ۲۶' ۲۴''	۲۲
۲	شادگان	۱۳۶۰-۱۳۹۱	۳۰° ۳۹' ۱۸''	۴۸° ۴۳' ۲۴''	۶
۳	چمن‌نظام	۱۳۷۰-۱۳۹۱	۳۰° ۴۴' ۳۳''	۴۹° ۵۵' ۵''	۱۷۸
۴	جوکنک	۱۳۷۰-۱۳۹۱	۳۱° ۲۰' ۲۸''	۴۹° ۴۳' ۲۵''	۳۰۶
۵	بهبهان	۱۳۷۰-۱۳۹۱	۳۰° ۴۰' ۲۹''	۵۰° ۲۰' ۳۲''	۱۷۸

<sup>۱</sup> Coruh River

<sup>۲</sup> Adaptive Neuro-Fuzzy Inference System (ANFIS)

<sup>۳</sup> Quinlan

<sup>۴</sup> Wang and Witten



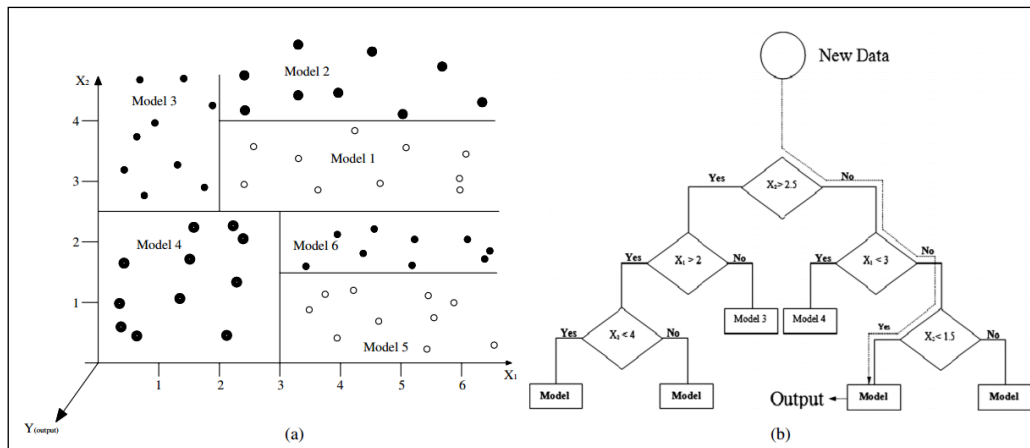
شکل ۲. موقعیت حوضه آبریز و ایستگاه‌های آب‌سنجی مورد مطالعه

که در آن،  $sd$  معرف انحراف معیار،  $T$  شامل نمونه‌هایی است که به گره مورد نظر رسیده‌اند و  $T_i$  شامل نمونه‌هایی است که از تقسیم گره مورد نظر بر اساس صفت انتخاب شده به دست آمده‌اند. الگوریتم  $M5$  تمامی حالت‌های مختلف جهت ایجاد شاخه بر اساس صفت خاص را بررسی کرده و در نهایت حالتی را انتخاب می‌کند که بتواند بیشتر از حالت‌های دیگر تابع خطای فوق را افزایش دهد. پس از تکمیل الگوریتم درختی برای نمونه‌های موجود در هر گره داخلی، یک مدل رگرسیون خطی چند متغیره برازش داده می‌شود. شکل ۲ نحوه تقسیم فضای مسئله به اجزای کوچک‌تر و کاربرد مدل‌های خطی را برای هر کدام از زیر دامنه‌ها با توجه به ساختار درختی نشان می‌دهد [۶].

در روش  $M5$  فضای مسئله به زیر دامنه‌هایی تقسیم شده و برای هر زیر دامنه یک مدل رگرسیون خطی چند متغیره برازش داده می‌شود. این الگوریتم؛ جداسازی‌های ممکن را در فضای چند متغیره انجام داده و به طور خودکار مدل‌هایی را برای هر کدام از دامنه‌ها ایجاد می‌کند [۱۱]. در این الگوریتم از پارامتر انحراف معیار مقادیر متغیر هدف به عنوان معیار اندازه‌گیری خطا جهت ایجاد شاخه در هر گره استفاده می‌شود. بدین صورت که صفتی که موجب کاهش بیشتر انحراف معیار در هر گره می‌شود به عنوان صفت مورد نظر جهت ایجاد شاخه انتخاب می‌شود. کاهش انحراف استاندارد<sup>۱</sup> که به عنوان تابع خطا در الگوریتم  $M5$  به کار می‌رود، به صورت زیر تعریف می‌شود:

$$SDR = sd(T) - \sum \left| \frac{T_i}{T} \right| \times sd(T) \quad (1)$$

<sup>۱</sup> Standard Deviation Reduction



شکل ۲. تقسیم فضای مسئله و ارائه مدل‌های خطی به ازای هر کدام از فضاها توسط الگوریتم M5

ساخت مدل‌های خطی از حالت لگاریتمی به توانی تبدیل شدند. برای مدل GP هم در ابتدا از لگاریتم طبیعی اعداد استفاده شد و پس از ساخت مدل، خروجی‌ها از حالت لگاریتمی به حالت توانی تبدیل شدند تا شرایط کاملاً یکسانی برای هر دو مدل داده‌کاوی فراهم شده باشد. همچنین داده‌های ورودی در مراحل آموزش و صحت‌سنجی برای دو مدل یکسان معرفی شدند. در این تحقیق از مجموع کل اطلاعات در دسترس ۸۰ درصد جهت آموزش مدل‌ها و ۲۰ درصد باقیمانده جهت صحت‌سنجی مورد استفاده قرار گرفت. جهت ساخت مدل‌های درختی با توجه به اطلاعات موجود از دبی جریان مربوط به ۴، ۳، ۲ و یک روز قبل و دبی جریان همان روز به‌عنوان پارامترهای ورودی به مدل‌ها معرفی شدند و نتایج مدل‌ها مورد بررسی قرار گرفت. نتایج اولیه نشان داد که تنها دبی جریان مربوط به یک روز قبل و دبی همان روز بر روی میزان دبی رسوب معلق تأثیر معنی‌داری داشته و به همین دلیل روابط ارائه شده در این تحقیق برای مدل‌های درختی بر اساس تنها دو پارامتر ورودی می‌باشند. در تحقیقات صورت گرفته از پارامترهای متعددی از قبیل دبی جریان همان روز و روزهای قبل، میزان بارندگی و دبی رسوب روزهای قبل استفاده شده است. در این تحقیق با توجه به برداشت ماهانه رسوب

برنامه‌ریزی ژنتیک (GP) از زیرشاخه‌های الگوریتم ژنتیک است که نخستین بار توسط کوزا<sup>۱</sup> معرفی شد [۹]. این روش از روش‌های الگوریتم گردشی<sup>۲</sup> به شمار می‌آید. تفاوت برنامه‌ریزی ژنتیک با الگوریتم ژنتیک در آن است که برنامه‌ریزی ژنتیک از ساختار درختی برای بهینه‌سازی مسئله استفاده می‌کند در حالی که الگوریتم ژنتیک بر روی رشته‌های بیتی کار می‌کند. در برنامه‌ریزی ژنتیک هیچ ساختار و ارتباطی میان متغیرهای ورودی و خروجی وجود ندارد و ساختار مدل و ضرایب بهینه نیز طی فرآیند بهینه‌سازی به دست می‌آیند. ساختار درختی از مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شود [۹]. سه عمل ژنتیکی تلاقی، جهش و تولیدمثل از جمله مهم‌ترین عمل‌های ژنتیکی در برنامه‌ریزی ژنتیک می‌باشند. عمل‌های دیگر از قبیل اصلاح ساختار با احتمال کمتری به کار گرفته می‌شوند. در تحقیق حاضر با توجه به اینکه مدل M5 تنها قادر به شبیه‌سازی مدل‌های خطی بوده و این امر با اساس معادلات متداول بار معلق رسوب در تناقض است، کلیه داده‌ها به صورت لگاریتم طبیعی به مدل M5 معرفی شدند و پس از

<sup>1</sup> Koza

<sup>2</sup> Circular algorithm

در روابط فوق  $Q_s^{measured}$  مقادیر مشاهداتی،  $Q_s^{computed}$  مقادیر پیش‌بینی شده،  $n$  تعداد داده‌ها و  $\overline{Q_s}$  میانگین مقادیر رسوب می‌باشند.

### ۳. نتایج

در این تحقیق، پس از معرفی داده‌های مشاهده‌ای به صورت لگاریتم طبیعی به مدل M5، ساختار درختی توسط این مدل برای ایستگاه‌های مورد مطالعه ایجاد شد که این ساختار در شکل ۳ ارائه شده است. همان‌گونه که اشاره شد مدل M5 تنها قادر به ساخت مدل‌های خطی بوده به همین دلیل و با توجه به مکانیزم انتقال رسوب که دارای رابطه‌ی توانی است، در این تحقیق کلیه مقادیر به صورت لگاریتمی به مدل معرفی گردید. مقادیر ارائه شده در شکل ۳ مقادیر واقعی بوده که از حالت لگاریتمی خارج شده‌اند.

شکل ۳ نشان می‌دهد که در تمامی ایستگاه‌ها به جز ایستگاه بهبهان، دبی جریان در همان روز به عنوان متغیر اصلی در ایجاد شاخه‌ها مورد استفاده قرار گرفته است. علاوه بر این در تمامی ایستگاه‌ها، مدل درختی جهت برآورد بار معلق تنها از دو شاخه استفاده کرده است و تنها در ایستگاه چمن‌نظام از چهار شاخه جهت محاسبه بار معلق استفاده شده است. مهم‌ترین دلیل این امر را می‌توان به مقادیر دبی و رسوب در این ایستگاه مربوط دانست که نسبت به بقیه ایستگاه‌ها دارای بازه تغییرات بیشتری است. معادلات ارائه شده توسط مدل درختی که در شکل ۳ با عنوان  $LM_i$  نشان داده شده است، معادلات خطی بوده که توسط الگوریتم درختی برای هر کدام از حالت‌ها توسعه داده شده است. این معادلات از حالت لگاریتمی خارج شده و به صورت معادلات ۵ تا ۱۶ تبدیل شده‌اند. معادلات ۵ تا ۸ مربوط به ایستگاه چمن‌نظام، معادلات ۹ و ۱۰ مربوط به ایستگاه بهبهان، معادلات ۱۱ و ۱۲ مربوط به ایستگاه جوکنک، معادلات ۱۳ و ۱۴ مربوط به ایستگاه شادگان و معادلات ۱۵ و ۱۶ مربوط به

معلق، استفاده از دبی رسوب روزهای قبل ممکن نبوده است. از طرفی هدف از این تحقیق ارائه مدلی ساده با کمترین متغیرهای ورودی به عنوان جایگزین روش منحنی سنج بوده، که با توجه به تحقیقات گذشته که در آن دبی جریان همان روز و روز قبل به عنوان مهم‌ترین پارامترهای مؤثر بر میزان بار معلق رسوب معرفی شده است [۱ و ۸]، تنها از این دو پارامتر استفاده گردید. در این تحقیق برای مدل‌سازی M5 و GP به ترتیب از برنامه‌های WEKA 3.7 و GeneXproTools 5.0 استفاده شد. در این تحقیق جهت مقایسه و ارزیابی کارایی مدل‌های ارائه شده از معیارهای آماری ریشه دوم میانگین مربعات خطا<sup>۱</sup> (RMSE)، ضریب تبیین<sup>۲</sup> ( $R^2$ )، نسبت اختلاف<sup>۳</sup> (DR) و ضریب ناش-ساتکلیف<sup>۴</sup> (NSE) استفاده شده است. ضریب NSE نشان‌دهنده مقدار نسبی واریانس باقیمانده در مقایسه با واریانس داده‌های مشاهداتی می‌باشد. چنانچه  $NSE > 0.7$  باشد، نتایج مدل‌سازی خوب بوده و مقادیر  $0.4 < NSE \leq 0.7$  رضایت‌بخش بوده ولی  $NSE \leq 0.4$  نشان‌دهنده دقت پایین مدل‌سازی خواهد بود [۱۴].

(۲)

$$R^2 = \frac{\left[ \sum (Q_s^{measured} - \overline{Q_s^{measured}})(Q_s^{computed} - \overline{Q_s^{computed}}) \right]^2}{\sum (Q_s^{measured} - \overline{Q_s^{measured}})^2 \sum (Q_s^{computed} - \overline{Q_s^{computed}})^2}$$

$$RMSE = \sqrt{\frac{1}{N} \sum (Q_s^{measured} - Q_s^{computed})^2}$$
 (۳)

$$D_R = \text{Log} \frac{Q_s^{computed}}{Q_s^{measured}}$$
 (۴)

$$NSE = 1 - \frac{\sum (Q_s^{measured} - Q_s^{computed})^2}{\sum (Q_s^{measured} - \overline{Q_s^{measured}})^2}$$
 (۵)

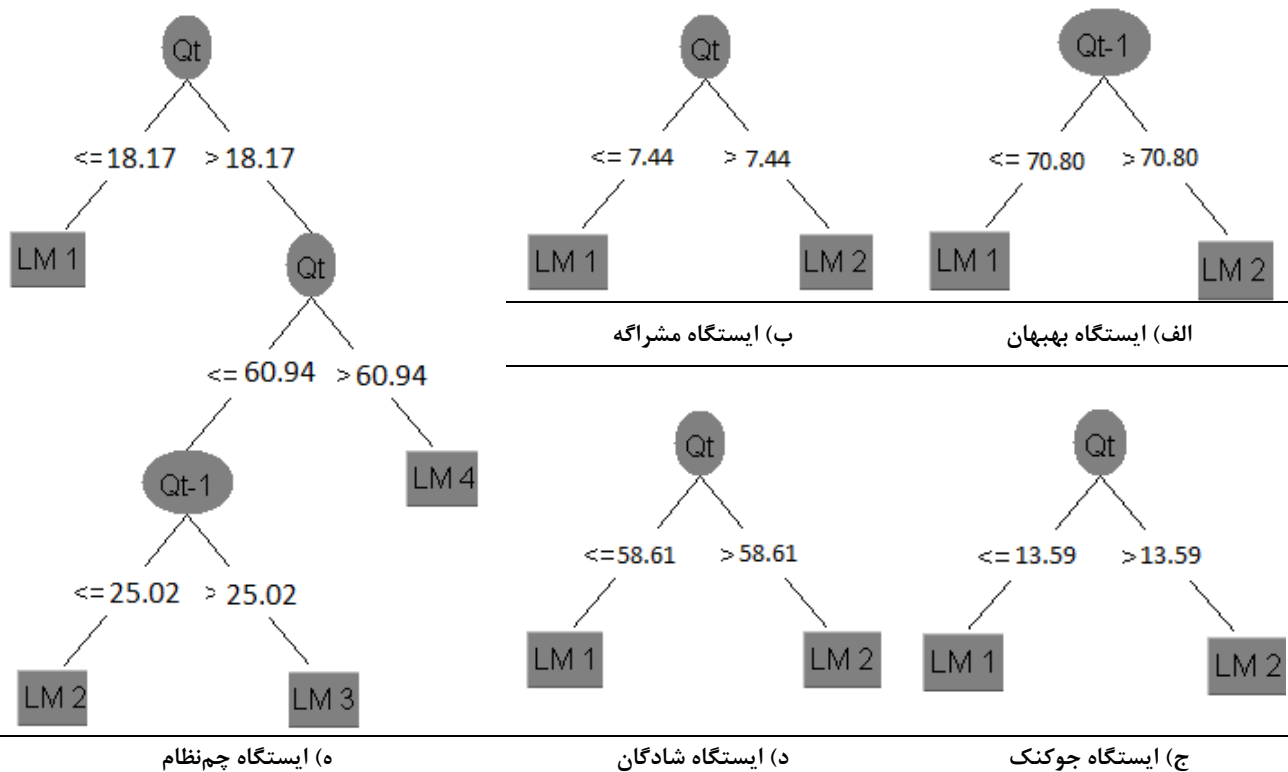
<sup>1</sup> Root Mean Square Error

<sup>2</sup> Determination coefficient

<sup>3</sup> Discrepancy ratio

<sup>4</sup> Nash-Sutcliffe coefficient

- (۱۱)  $LM1: Q_s = 4.9Q_{t-1}^{0.95} Q_t^{1.37}$  ایستگاه مشراکه است. در این روابط  $Q_s$  دبی رسوب و  $Q_t$  و  $Q_{t-1}$  دبی جریان همان روز و یک روز قبل است.
- (۱۲)  $LM2: Q_s = 0.57Q_t^{2.17}$   $LM1: Q_s = 5.81Q_{t-1}^{0.38} Q_t^{0.44}$  (۵)
- (۱۳)  $LM1: Q_s = 16.11Q_t^{0.6}$   $LM2: Q_s = 487.84Q_{t-1}^{1.16} Q_t^{0.82}$  (۶)
- (۱۴)  $LM2: Q_s = 1.5Q_t^{2.29}$   $LM3: Q_s = 1.07Q_{t-1}^{0.68} Q_t^{0.6}$  (۷)
- (۱۵)  $LM1: Q_s = 2.27Q_{t-1}^{1.4}$   $LM4: Q_s = 0.004Q_{t-1}^{0.35} Q_t^{2.98}$  (۸)
- (۱۶)  $LM2: Q_s = 0.15Q_t^{2.18}$   $LM1: Q_s = 5.002Q_{t-1}^{0.83} Q_t^{0.074}$  (۹)
- $LM2: Q_s = 0.88Q_t^{1.45}$  (۱۰)



شکل ۳. ساختار درختی M5 جهت برآورد بار معلق رسوب در ایستگاه‌های مختلف

قبل نسبت به دبی همان روز کاهش داشته و یا این پارامتر از معادلات ارائه شده حذف گردیده است.

در مدل GP با افزودن عملگرهای  $\tan h$ ,  $\text{power}$ ,  $\text{sqrt}$ ,  $\tan^{-1}$ ,  $\text{arccot}$  و  $\text{invers. arctan}$ ,  $\text{cot}^{-1}$ ,  $\text{coth}$ ,  $\text{cot}$ ,  $\text{tan}$ ,  $\text{tan}^{-1}$  چهار عملگر اصلی جمع، تفریق، ضرب و تقسیم، نتایج بهتری حاصل نگردید. بر همین اساس جهت ساخت مدل‌ها تنها از چهار عملگر اصلی استفاده گردید تا پیچیدگی معادلات ارائه شده نیز کمتر گردد. معادلات ۱۷ تا ۲۱ به‌عنوان بهترین و دقیق‌ترین مدل‌های ارائه شده توسط مدل GP به‌ترتیب برای ایستگاه‌های شادگان، مشراکه، جوکنک، بهبهان و چمن نظام انتخاب شدند. شکل ۴ نشان‌دهنده ساختار درختی مدل GP جهت برآورد بار معلق رسوب در ایستگاه شادگان بوده که حاصل جمع سه درخت ارائه شده در این شکل، مدل نهایی را ارائه می‌دهد.

در تمامی معادلات ارائه شده توسط مدل درختی (معادلات ۵ تا ۱۶)، میان دبی رسوب و دبی جریان مربوط به همان روز و روز قبل رابطه مستقیمی وجود داشته که با فیزیک مسئله انتقال رسوب مطابقت دارد. در ایستگاه چمن نظام که بیشترین شاخه‌ها را در میان بقیه ایستگاه‌ها دارد، چهار معادله ارائه شده است که در سه معادله ۵ تا ۷، توان مربوط به دبی همان روز و روز قبل تفاوت چندانی با هم ندارند که نشان‌دهنده تأثیر نسبتاً یکسان این دو متغیر بر روی میزان غلظت رسوب معلق است. ولی در معادله ۸ توان مربوط به دبی همان روز به طور محسوسی نسبت به دبی روز قبل افزایش یافته است. با توجه به ساختار درختی ارائه شده در شکل ۳، معادله ۸ مربوط به دبی‌های بالای ۶۰ مترمکعب بر ثانیه بوده که به‌عنوان دبی سیلابی در این ایستگاه به حساب می‌آیند. همین روند در مورد بقیه ایستگاه‌ها نیز صادق است. به‌گونه‌ای که در دبی‌های بالا، توان مربوط به دبی روز

$$Q_s = (Q_{t-1} (3.14-3.54) + (Q_t+2.7)) + ((1.56-Q_t) \left( \frac{Q_{t-1}}{-3.6} - \frac{Q_t-Q_{t-1}}{-3.6+Q_t} \right)) + \left( \frac{Q_{t-1}(Q_{t-1}-5/52)}{9.92+Q_t} \right) + \left( \frac{-2.7}{Q_t} \right) \quad (17)$$

$$Q_s = \left( \frac{Q_{t-1}-Q_t}{0.2+(2Q_{t-1})} \right) + \left( \frac{Q_{t-1}-0.30}{2Q_{t-1}} \right) \left( \frac{Q_t}{5.54} \right) Q_{t-1} + \left( \frac{53.74}{5.54(4.8-\frac{Q_{t-1}}{Q_t})} \right) \quad (18)$$

$$Q_s = \left( \frac{-46.80-Q_{t-1}}{Q_{t-1}-46.80} \right) - \left( \frac{Q_t+4.5}{-46.80Q_t} \right) + \left( \frac{-5.5(\frac{3.3}{Q_{t-1}}-0.90)}{-9.4(-46.80Q_t)} \right) + \left( \frac{(Q_t+5.3)-Q_t^2}{\frac{-11.81}{3.4+Q_t}} \right) \quad (19)$$

$$Q_s = \left( \frac{Q_t+Q_{t-1}}{Q_{t-1}^2+(1.13-Q_t)} \right) + ((0.086) (Q_t (Q_{t-1} - 1.2))) + (Q_{t-1} + \left( \frac{5.4Q_t}{3965.89(4.07-Q_t)} \right)) \quad (20)$$

$$Q_s = \left( \frac{Q_{t-1}(-3.6Q_{t-1})(2.8-Q_t)}{60.74} \right) + \left( \frac{(-11.11-Q_t)}{Q_{t-1}^2(2.8Q_{t-1})} \right) + \left( \frac{Q_t}{Q_{t-1}} - Q_t \right) + \left( \frac{Q_{t-1}}{Q_{t-1}^2} \right) \quad (21)$$

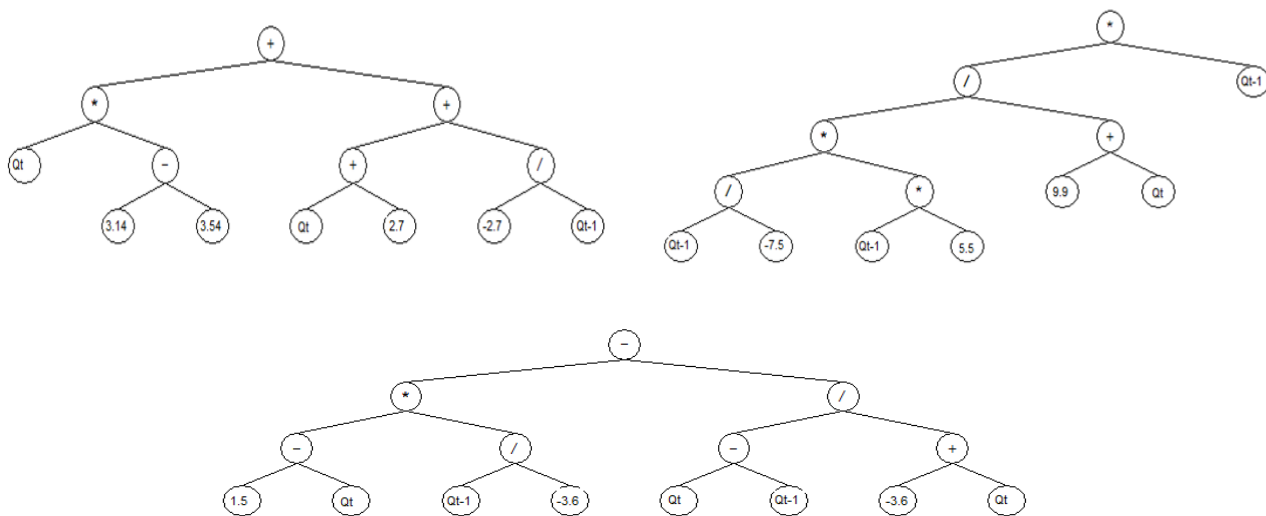
ارائه شده؛ محور افقی نشان‌دهنده مقادیر رسوب معلق مشاهده شده و محور قائم نیز معرف مقادیر محاسبه شده توسط الگوریتم‌های درختی و منحنی سنجه رسوب است. بر اساس گراف‌های ارائه شده هر چه پراکندگی داده‌ها در

نتایج مربوط به مقادیر بار معلق محاسبه شده توسط مدل‌های درختی  $GP$ ،  $M5$  و منحنی سنجه رسوب در ایستگاه‌های آب‌سنجی مطالعه شده در این تحقیق در شکل‌های ۵ تا ۷ نشان داده شده است. در کلیه گراف‌های

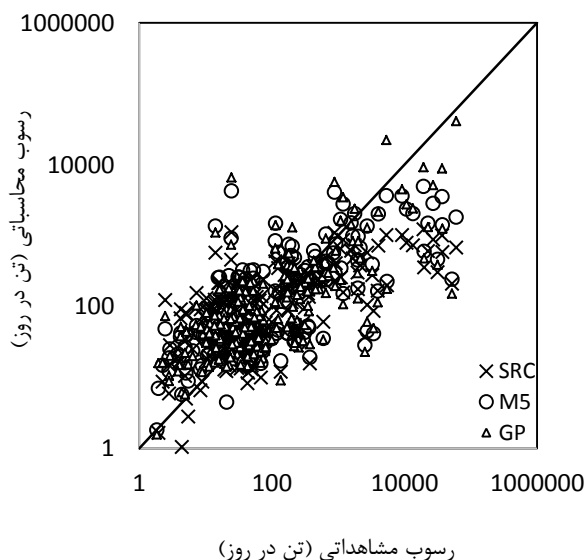
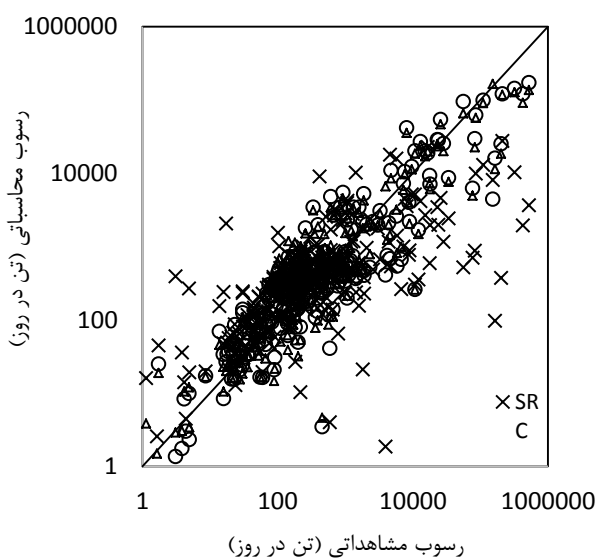


مقدار رسوب معلق را نسبت به الگوریتم‌های درختی کمتر از مقدار واقعی برآورد نموده است. این امر با توجه به اهمیت بار معلق رسوب در دبی‌های سیلابی و در نظر گرفتن ضریب اطمینان جهت پروژه‌های مهندسی رودخانه می‌تواند به‌عنوان یکی از نقاط ضعف روش منحنی سنجه رسوب به حساب آید.

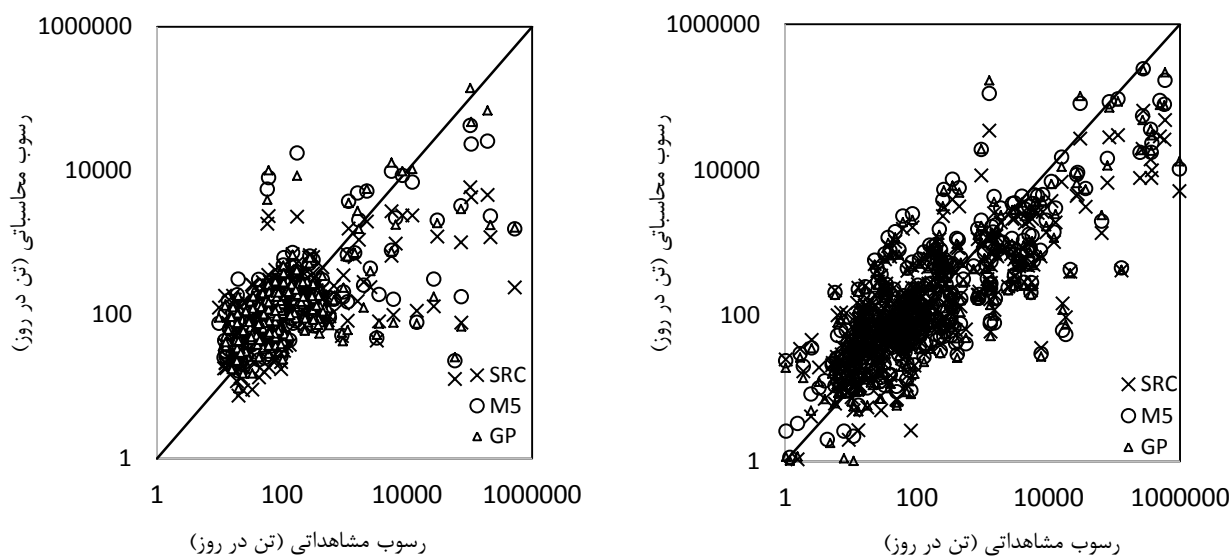
اطراف خط ۴۵ درجه بیشتر باشد، نشان‌دهنده دقت بالاتر مدل خواهد بود. در کلیه گراف‌ها، نتایج مدل‌های M5 و GP در مقایسه با منحنی سنجه انطباق بهتری با مقادیر مشاهداتی داشته و پراکندگی بهتری در اطراف خط ۴۵ درجه نشان می‌دهند. علاوه بر این، روش منحنی سنجه که متداول‌ترین روش محاسبه بار معلق به حساب می‌آید، در دبی‌های سیلابی که حداکثر رسوب را به همراه دارند،



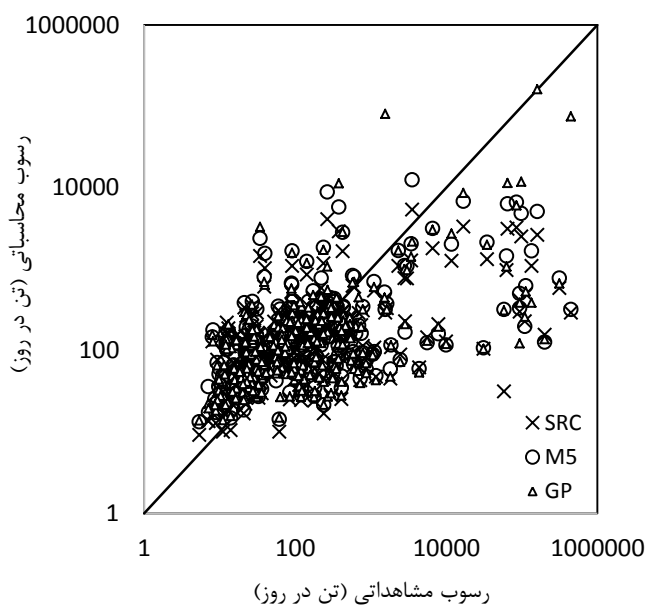
شکل ۴. ساختار درختی GP جهت برآورد بار معلق رسوب معلق در ایستگاه شادگان



شکل ۵. مقایسه پراکندگی نتایج مدل‌های درختی و منحنی سنجه رسوب در ایستگاه‌های شادگان و مشراکه



شکل ۶. مقایسه پراکندگی نتایج مدل‌های درختی و منحنی سنجه رسوب در ایستگاه‌های جوکنک و چمن‌نظام



شکل ۷. مقایسه پراکندگی نتایج مدل‌های درختی و منحنی سنجه رسوب در ایستگاه بهبهان

کامل نتایج مدل و مقادیر واقعی است. در صورتی که  $(DR > 0)$  باشد مدل مورد استفاده، بار معلق رسوب را بیشتر از مقدار واقعی برآورد کرده است. در غیر این صورت  $(DR < 0)$ ، بار معلق رسوب کمتر از مقدار واقعی محاسبه شده است. در جدول ۲ درصد مقادیر با DR بین

جهت بررسی دقت مدل‌ها در این تحقیق از شاخص‌های آماری متنوعی استفاده گردید که نتایج تجزیه و تحلیل‌های آماری صورت گرفته در جدول ۲ ارائه شده است. یکی از مهم‌ترین شاخص‌های مورد استفاده، DR است. چنانچه DR برابر صفر باشد نشان دهنده تطابق

درختی و منحنی سنجه رسوب در شکل‌های ۸ تا ۱۲ نشان داده شده است.

۰/۳- و ۰/۳+ به‌عنوان دقت هر مدل در نظر گرفته شده است. علاوه بر این هیستوگرام فراوانی DR برای دو مدل

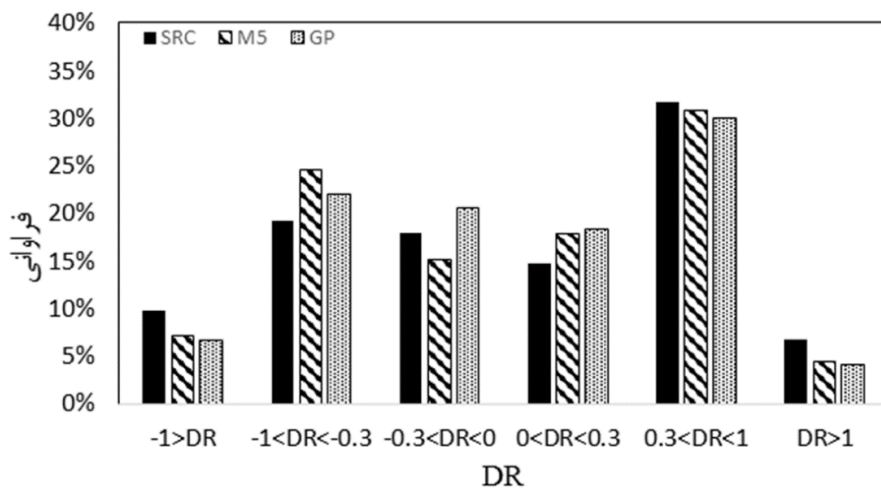
جدول ۲. بررسی کارایی مدل‌های هوشمند و تجربی در تخمین بار معلق رسوب

ایستگاه	مدل	DR<-۰/۳	-۰/۳<DR<۰	۰<DR<۰/۳	DR>۰/۳	دقت	R <sup>2</sup>	RMSE	NSE
شادگان	مدل M5	۲۲	۲۱	۱۸	۳۰	۳۹	۰/۵۰	۰/۶۹	۰/۵۹
	مدل GP	۲۵	۱۵	۱۸	۳۱	۳۳	۰/۴۹	۰/۶۹	۰/۲۵
	منحنی سنجه رسوب	۱۹	۱۸	۱۵	۳۲	۳۳	۰/۳۸	۰/۷۷	۰/۴۸
مشراکه	مدل M5	۱۹	۲۶	۲۵	۲۶	۵۱	۰/۸۱	۰/۴۶	۰/۵۱
	مدل GP	۱۸	۲۸	۲۲	۳۰	۵۰	۰/۷۳	۰/۵۶	۰/۷۰
	منحنی سنجه رسوب	۱۵	۱۶	۲۵	۳۴	۴۰	۰/۴۴	۰/۷۸	۰/۳۱
چوکنک	مدل M5	۲۴	۱۸	۲۱	۲۵	۳۸	۰/۶۹	۰/۷۰	۰/۴۷
	مدل GP	۲۲	۱۸	۲۰	۲۵	۳۸	۰/۶۷	۰/۶۹	۰/۴۸
	منحنی سنجه رسوب	۲۱	۱۶	۱۸	۲۸	۳۴	۰/۶۳	۰/۷۶	۰/۴۲
چمن‌نظام	مدل M5	۱۷	۲۰	۳۱	۲۴	۵۲	۰/۴۵	۰/۶۳	۰/۴۷
	مدل GP	۱۵	۱۸	۳۲	۲۷	۵۰	۰/۴۱	۰/۶۶	۰/۶۹
	منحنی سنجه رسوب	۱۷	۱۵	۲۶	۳۲	۴۲	۰/۳۱	۰/۷۱	۰/۳۴
بهبهان	مدل M5	۱۵	۱۲	۲۷	۳۳	۳۹	۰/۳۵	۰/۷۲	۰/۵۵
	مدل GP	۱۵	۱۰	۲۷	۳۴	۳۷	۰/۳۴	۰/۷۳	۰/۳۰
	منحنی سنجه رسوب	۱۵	۹	۲۵	۳۷	۳۴	۰/۲۸	۰/۷۷	۰/۵۱

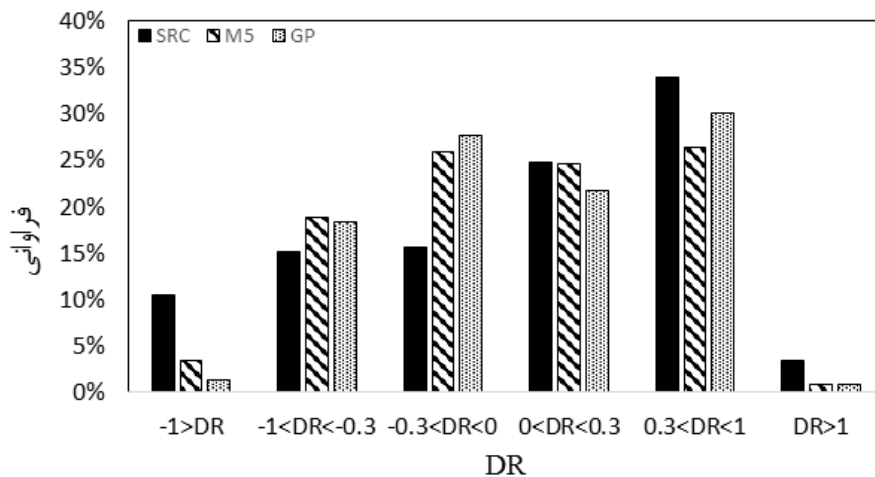
رسوب در دو ایستگاه دارای NSE کمتر از ۰/۴ بوده که نشان‌دهنده نامطلوب بودن نتایج این دو روش در برآورد رسوب معلق در آن ایستگاه‌ها می‌باشد.

با توجه به مقادیر DR، در تمامی ایستگاه‌ها دو مدل GP و M5 نسبت به منحنی سنجه رسوب دارای دقت بالاتری می‌باشند. جهت بررسی دقیق‌تر پارامتر DR در ایستگاه‌های مختلف، مقادیر این پارامتر به‌صورت هیستوگرام فراوانی در شکل‌های ۸ تا ۱۲ ارائه شده است. در هیستوگرام‌های ارائه شده،  $DR < -1$  نشان‌دهنده مقادیری است که توسط روش مورد نظر، بیش از ۱۰ برابر کمتر از مقادیر مشاهداتی محاسبه شده است و در  $DR > 1$ ، بار معلق رسوب بیش از ۱۰ برابر بیشتر از مقادیر مشاهداتی برآورد شده است.

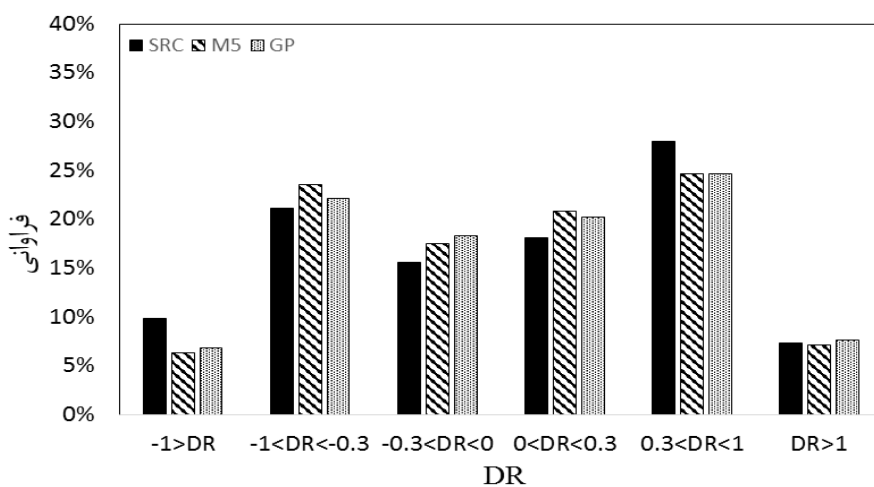
نتایج جدول ۲ نشان می‌دهد که در مجموع مدل‌های داده‌کاوی بار رسوب معلق را با خطای کمتر و همبستگی بالاتر برآورد می‌کنند. در جدول ۲ ضریب تبیین برای مدل‌های داده‌کاوی نسبت به منحنی سنجه رسوب به ۱ نزدیک‌تر بوده و در بعضی ایستگاه‌ها از قبیل ایستگاه مشراکه تفاوت زیادی میان مقادیر این ضریب در مدل‌های داده‌کاوی و منحنی سنجه مشاهده می‌شود. در مورد ریشه دوم میانگین مربعات خطا نیز همین روند مشاهده می‌شود به‌گونه‌ای که در ایستگاه مشراکه ریشه دوم میانگین مربعات خطای منحنی سنجه رسوب ۱/۵ برابر مدل درختی M5 محاسبه شده است. مقادیر ضریب ناش-ساتکلیف مدل درختی M5 در تمامی ایستگاه‌های آب‌سنجی بیشتر از ۰/۴ بوده که نشان‌دهنده رضایت‌بخش بودن نتایج این مدل می‌باشد. مدل GP و منحنی سنجه



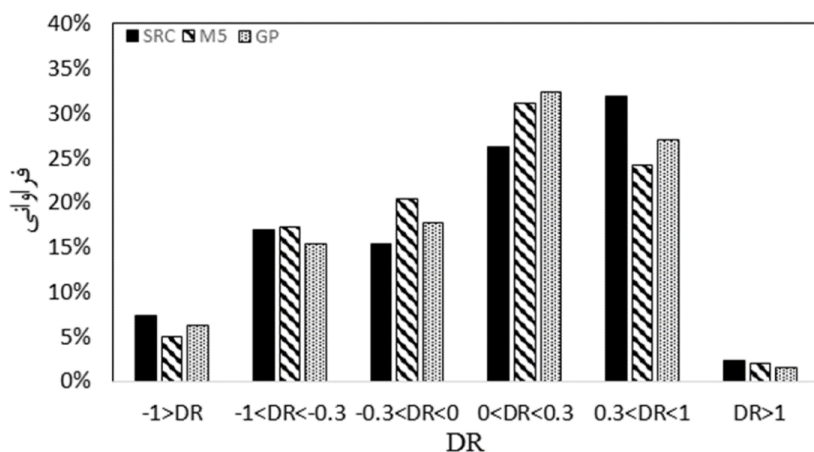
شکل ۸. مقایسه نسبت اختلاف ( $DR$ ) روش‌های مختلف در ایستگاه شادگان



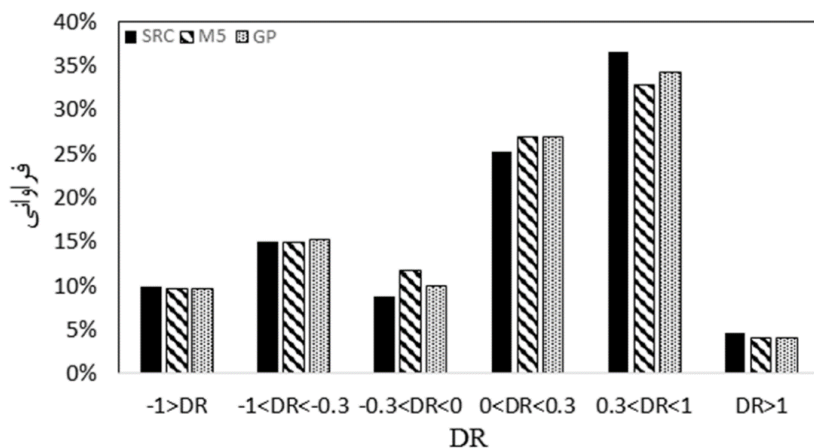
شکل ۹. مقایسه نسبت اختلاف ( $DR$ ) روش‌های مختلف در ایستگاه مشراکه



شکل ۱۰. مقایسه نسبت اختلاف ( $DR$ ) روش‌های مختلف در ایستگاه جوکنک



شکل ۱۱. مقایسه نسبت اختلاف ( $DR$ ) روش‌های مختلف در ایستگاه چم‌نظام



شکل ۱۲. مقایسه نسبت اختلاف ( $DR$ ) روش‌های مختلف در ایستگاه بهبهان

می‌دهد که نتایج دو مدل درختی نسبت به منحنی سنجه رسوب از دقت بالاتری برخوردار می‌باشند. ایستگاه مشراکه بیشترین تفاوت را میان نتایج مدل‌های درختی و منحنی سنجه داشته است، به گونه‌ای که ریشه دوم میانگین مربعات خطا جهت مدل‌های M5 و GP به ترتیب به میزان ۳۲ و ۲۲ درصد بهبود یافته است. در ایستگاه بهبهان نیز کمترین تفاوت میان مدل‌های درختی و منحنی سنجه مشاهده گردید که این اختلاف بین ۴ تا ۵ درصد برآورد گردید. مقادیر ضریب ناش-ساتکلیف مدل درختی M5 در تمامی ایستگاه‌های آب‌سنجی از ۰/۴

#### ۴. بحث و نتیجه‌گیری

در این تحقیق سعی گردید با به‌کارگیری الگوریتم‌های M5 و GP و تنها با استفاده از دو پارامتر دبی جریان روز قبل و همان روز اقدام به برآورد بار معلق رسوب گردد. علاوه بر این از روش منحنی سنجه رسوب که روش متداول محاسبه بار معلق در رودخانه‌ها به حساب می‌آید، جهت بررسی دقت مدل‌های درختی استفاده گردید. تجزیه و تحلیل‌های آماری از قبیل ریشه دوم میانگین مربعات خطا و ضریب ناش-ساتکلیف، نشان

ایستگاه‌ها به ویژه در ایستگاه‌های مشراکه و چم‌نظام به وضوح مشاهده می‌گردد. نتایج تحقیق صورت گرفته توسط [۱] که جهت شبیه‌سازی رسوب معلق از پارامترهای ورودی متعددی استفاده کرده است، نشان می‌دهد که متغیرهای دبی جریان با فاصله زمانی بیشتر از یک روز تأثیر معنی‌داری بر روی رسوب معلق ندارد، که در این تحقیق نیز تنها از دو پارامتر دبی جریان همان روز و روز قبل به‌عنوان پارامترهای ورودی استفاده گردید. علاوه بر این مطالعه صورت گرفته توسط [۸] نشان می‌دهد، پارامتر دبی جریان همان روز و پس از آن دبی جریان روز قبل به‌عنوان تأثیرگذارترین پارامترها بر روی دبی رسوب معلق می‌باشند. در این تحقیق نیز در اغلب مدل‌های درختی، پارامتر دبی جریان همان روز تأثیر بیشتری نسبت به دبی جریان روز قبل داشته است. در نهایت مقایسه کارایی دو مدل درختی نشان داد که نتایج دو مدل نزدیک به هم بوده اما روابط پیشنهادی توسط مدل GP به‌مراتب پیچیده‌تر از روابط مدل M5 است. بر این اساس می‌توان مدل M5 را به‌عنوان مدل پیشنهادی با توجه به دقت بالا و سادگی معادلات ارائه شده و قابل درک بودن آن جهت برآورد بار معلق رسوب پیشنهاد کرد.

بیشتر بوده که نشان‌دهنده مطلوب بودن نتایج این مدل در مقایسه با دو روش GP و منحنی سنجه رسوب می‌باشد. علاوه بر این بر اساس هیستوگرام فراوانی نسبت اختلاف (DR)، منحنی سنجه رسوب بیشترین مقادیر با  $DR < -1$  که نشان‌دهنده برآورد بسیار پایین مقدار رسوب معلق است، را دارا است. همین نکته در مورد  $DR > 1$  نیز وجود دارد، بدین معنی که منحنی سنجه رسوب درصد بیشتری از مقادیر بار معلق را با اختلاف زیاد بیشتر از مقادیر واقعی برآورد می‌کند. بر این اساس روش منحنی سنجه رسوب نسبت به دو روش دیگر دارای درصد برآوردهای با دقت پایین بیشتری در تخمین مقدار رسوب معلق بوده که این نکته می‌تواند به‌عنوان مهم‌ترین ضعف این روش به حساب آید. نتایج این تحقیق مانند مطالعه صورت گرفته توسط [۱۲] نشان می‌دهد که در دبی‌های بالا دقت روش منحنی سنجه رسوب به شدت کاهش می‌یابد، این در حالی است که مدل درختی در این شرایط مقدار رسوب را با دقت بالاتری شبیه‌سازی کرده است. این نکته از این نظر حائز اهمیت است که در پروژه‌های مهندسی رودخانه، شرایط سیلابی از اهمیت بیشتری برخوردار بوده و مدل‌های درختی در این شرایط از کارایی بالاتری برخوردار می‌باشند. این امر در تمامی

## References

- [1] Buyukyildiz, M. and Kumcu, S.Y. (2017). An estimation of the suspended sediment load using adaptive network based fuzzy inference system, support vector machine and artificial neural network models. *Water Resources Management*, 31, 1343-1359.
- [2] Dehghani, A.A., Zanganeh, M.E., Mosaede, A. and Kohestane, N. (2010). Neural network and sediment rating curve methods (Case Study: Doogb River in Golestan Province). *Journal of Agricultural Sciences and Natural Resources*, 16, 1-12 (In Persian).
- [3] Ebrahimi Mohamadi, S.h., Sadeghi, H. and Chapi, K. (2013). Analysis of runoff, nutrient-rich sediment input to the lake at different times of storm and ground water. *Journal of Soil and Water Conservation*, 2, 61-74 (In Persian).
- [4] Etemad-Shahidi, A. and Ghaemi, N. (2011). Model tree approach for prediction of pile groups scour due to waves. *Journal of Ocean Engineering*, 38, 1522-1527.

- [5] Etemad-Shahidi, A. and Mahjoobi, J. (2009). Comparison between M5' model tree and neural networks for prediction of significant wave height. *Journal of Ocean Engineering*, 36(15-16): 1175-1181.
- [6] Etemad-Shahidi, A. and Taghipour, M. (2012). Predicting longitudinal dispersion coefficient in natural streams using M5' model tree. *Journal of Hydraulic Engineering*, 138(6): 542-554.
- [7] Jabari, A. and Samadi, M. (2014). M5 algorithm to predict scour depth downstream of overflow. *Applied Research*. Iran Water Resources Management Company (In Persian).
- [8] Kakaei Lafdani, E., Moghaddam Nia, A. and Ahmadi, A. (2013). Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Journal of Hydrology*, 478, 50-62.
- [9] Koza, J. R. (1992). Genetic Programming: on the programming of computers by means of natural selection. Cambridge, MA: MIT Press.
- [10] Mahjoobi, J., Sabzianpoor, A. and Jabbari, E. (2010). Application of Meta-Heuristic models for local scour evaluation. Ninth International Conference on Computing Anticipatory Systems, AIP Conference Proceedings, 1303, 389-397.
- [11] Quinlan, J. R. (1992). Learning with continuous classes. Proc., 5th Australian Joint Conf. on Artificial Intelligence, World Scientific, Singapore, 343-348.
- [12] Talebi, A., Mahjoobi, J., Dastorani, M.T. and Moosavi, V. (2017). Estimation of suspended sediment load using regression trees and model trees approaches (Case study: Hyderabad drainage basin in Iran). *Journal of Hydraulic Engineering*, 23, 212-219.
- [13] Wang, Y. and Witten, I. H. (1997). Induction of model trees for predicting continuous classes. Proc. of the Poster Papers of the European Conference on Machine Learning, University of Economics, Faculty of Informatics and Statistics, Prague, Czech Republic.
- [14] Wu, B., Wang, Z., Zhang, Q., Shen, N. and Liu, J. (2017). Modelling sheet erosion on steep slopes in the loess region of China. *Journal of Hydrology*, 553, 549-558.
- [15] Yang, C.T., Marsooli, R. and Aalami, M.T. (2009). Evaluation of total load sediment transport formulas using ANN. *International Journal of Sediment Research*, 24 (3), 274-286.
- [16] Zahiri, J. (2016). Application of nonparametric CART and M5' methods on bridge piers scour evaluation. *Journal of Irrigation and Water Engineering*, 20, 50-35 (In Persian).
- [17] Zhang, J., Zhang, X., Li, R., Chen, L. and Lin, P. (2017). Did stream flow or suspended sediment concentration changes reduce sediment load in the middle reaches of the Yellow River? *Journal of Hydrology*, 546, 357-369.

